Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 10, No. 4, December 2022, pp. 825~836
ISSN: 2089-3272, DOI: 10.52549/ijeei.v10i4.3946

Forecasting and Clustering of Cassava Price by Machine Learning (A study of Cassava prices in Thailand)

Sayan Tepdang[1], Ratthakorn Ponprasert[2]
[1,2]Faculty of Business Administration and Information Technology, Rajamangala University of Technology Tawan-Ok, Chakrabongse Bhuvanarth Campus, Bangkok, Thailand

Article Info

Article history:

Received Jun 23, 2022
Revised Nov 15, 2022
Accepted Nov 28, 2022

Keywords:

Machine Learning
Data Mining
Forecasting of Cassava Price
Data Analysis
Knowledge Discovery in Data

ABSTRACT

Forecasting and clustering the price of cassava is essential for agriculture, but the difficult part of forecasting is price fluctuation, in which the fact of prices is going up and down and be changed monthly. The paper proposes to forecast the price of Cassava by machine learning. The process had been calculated by the price of Cassava from January 2005 to February 2022, which has been collected for 17 years by the Office of Agricultural Economics, Ministry of Agriculture, and Cooperatives. The research on forecasting found that the method of Support Vector Machine including using add-on feature with Garlic Price and Potato Price showing the Root Mean Squared Error (RMSE) with the lowest point as of 0.10. If comparing to the Conventional method with the equal database. The result shows that the proposed method demonstrates the value of the Mean Absolute Percentage Error (MAPE) as 3.35%, it displays more effectively as 0.61%. For the final process of clustering the price by analyzing with K-mean, the result came up with a peak pricing period in December of 14.08%. Subsequently, the agricultures would apply the research result to implement their planting plan for profit-making.

Corresponding Author:

Sayan Tepdang
Faculty of Business Administration and Information Technology, Rajamangala University of Technology Tawan-Ok, Chakrabongse Bhuvanarth Campus, Bangkok, Thailand
Email: sayan_te@rmutto.ac.

1. INTRODUCTION

Cassava is plant-based, which is a high-ranking in the top five subordinating to Corn, Wheat, Glutinous Rice, and Potato. It also has a high quantity of carbohydrate as well as provide various benefits. For instance, the food - as the consumable product for human and husbandry, the fuel – as the main ingredient for the Ethanol production process, and the manufacturing – as the component of the Alcohol and Apparel production process, etc. In 2019, the global production of Cassava came up to 309 million tons. Africa is the largest source of global products leading to Asia, America, and Oceania continent. For the country, Nigeria is the top producer leading to Congo, Thailand, Ghana, Brazil, and Indonesia. According to Thailand by 2020, the country has a cultivated land of approximately 3.5 million acres located in Nakhonratchasima, Kampangpetch, Chaiyaphum, and Kanchanaburi, respectively[1]. The cost of planting cassava in Thailand is approximately 5,974.78 THB for 0.39 acre. The average cost from 2011 to 2021 is 3.398 tons for 0.39 acre [2].

According to the methodology of the research, there are 2 approaches; 1) the Mathematical and Statistical approach and 2) the Machine Learning approach. Firstly, the Autoregressive Integrated Moving Average (ARIMA) method in terms of the Mathematic and Statistic, which is a showcase of predication, is implied in the related research papers to forecast the price of agricultural products. For instance, aromatic coconut [3], field crops [4], litopenaeus vannamei [5], and mango export volumes [6]. In addition, the method also implied forecasting for the Cassava product, for example: fresh cassava root buying prices and cassava chip selling prices [7], export volumes of cassava [8], and cassava price [9] ,[10]. As a result, the method is a fast approach for processing and forecasting, meanwhile, both variable and suitable methods remain an important part to provide the study result precisely. Secondly, the Machine Learning approach, which refers to Artificial Neural Network, k-Nearest Neighbor, Linear Regression Analysis, Deep Learning Support Vector Machine, etc. There are research papers, which are implied by these methods, for instance, stock market [11] ,[24] , [25], selling [12], finance [13], relating usage in forecasting growing cassava [14], or the price of cassava [15], [16] Moreover, the methods would be a suitable approach depending on the add-on features and the condition of variables, which are implied in forecasting the cassava price.

In summary, the previous research papers demonstrate that Machine Learning might be a suitable approach to forecasting the price of cassava because the approach provides various ways in the analytical process. As a result of forecasting, the study requires more add-on features combing with the forecasting process for an effective outcome, as well as, acknowledging the high season period that would be costed with the best price of cassava including the planting plan for the cultivalors who would earn high income for this product with the best price.

2. THE ANALYSIS OF CASSAVA PRICE

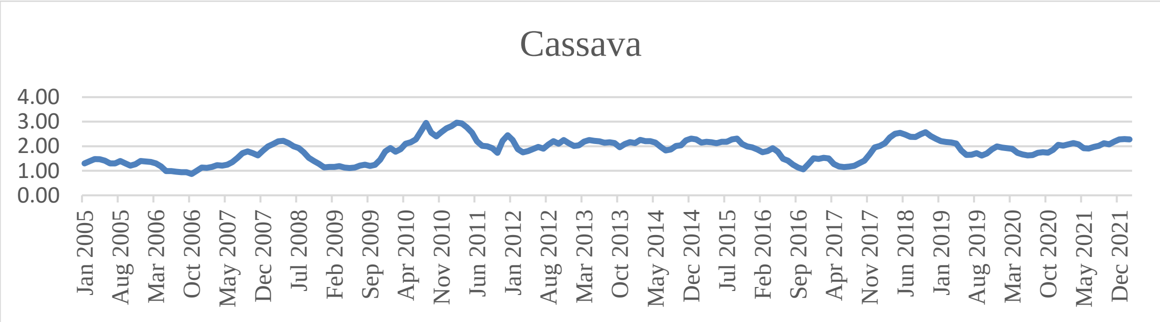


Figure 1. The price of cassava during January 2005 – February 2022

In Regard to Figure 1, the prices of cassava from January 2005 to February 2022 demonstrate that the sequence of cassava's price becomes a part of the marketing trend-stationary, which directly makes the sequence unsteady. The cassava price is 1.84 THB/KG, the highest price is 2.96 THB/KG and the lowest price is 0.87 THB/KG even if compared with other agricultural products such as Garlic and Potato, the market price also presents a similar trend as below graph.

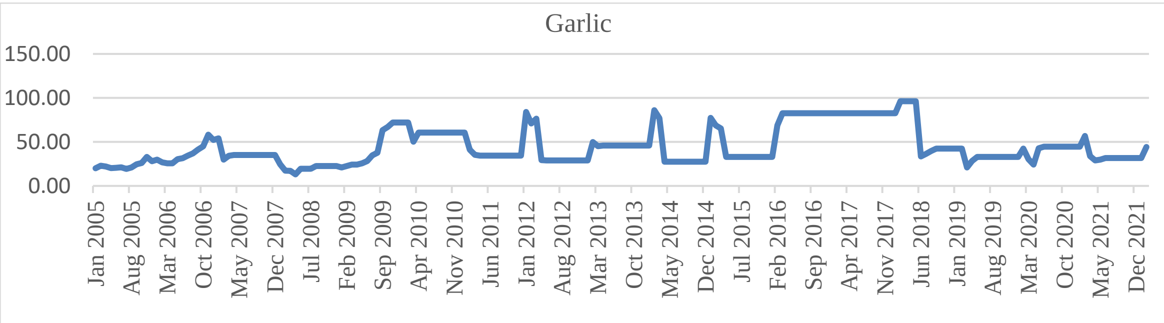


Figure 2. Garlic Price from January 2005 – February 2022

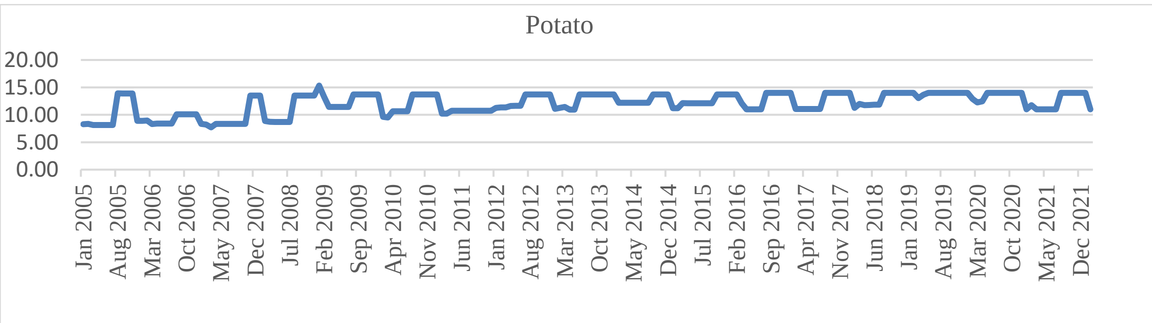


Figure 3. Potato Price from January 2005 – February 2022

3. RESEARCH METHOD

The research process as figure 4, which is implied by the Cross-Industry Standard Process for Data Mining or CRISP-DM, mainly is a conceptual framework for the forecasting and clustering process with the 6 processes as the following details [17] [18].

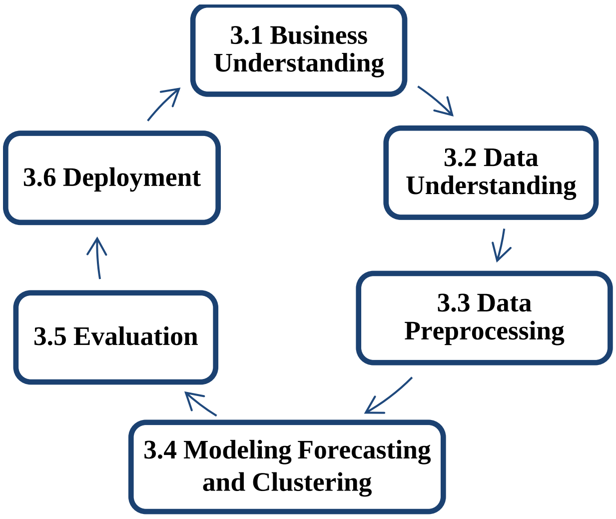


Figure 4. Research Process

3.1 Business Understanding

To understand the cassava agriculturists, planting period, investment, and amount of agricultural product in order to ascertain the price of cassava with relating agricultural products through the forecasting and clustering process for the cassava price.

3.2 Data Understanding

To acknowledge the database of cassava's price for 17 years from January 2005 to February 2022, which is collected by the Office of Agricultural Economics, Ministry of Agriculture and Cooperatives [19]. Then, finding the suitable approach for the forecasting process as well as choosing the agricultural product with 11 types such as Mung Bean, Shallot, Garlic, Onion, Soy, Maize, Rice, Sugar Cane, Palm Kernel, Oil, and Coffee for comparing in the analytical process.

3.3 Data Preprocessing

Preparing database to process as follows- 1) To consider and analyze the Correlation and Association Rules 2) To forecast, and 3) To find the best period of the highest cost of cassava. The data has been set for analysis was generated as 2 types; (1) The first type of data are implied to the process of forecasting, finding correlation, and price clustering, these type of data were arranged in a time-series basis during January 2005 to February 2022, which is shown in figure 5, and (2) The second type of data are implied for considering based on the association rules representing the relation between the price of Cassava and the others agriculture products, which was converted to value True or False, the outcome of this process is shown as an example of converting process is shown in figure 6.

Mung Bean	Shallot	Garlic	Potato	Onion	Soy	Maize	Rice	Sugarcane	Palm Kernel Oil	Coffee	Before Cassava	Cassava
15.62	10.12	20.00	8.28	5.90	13.00	5.01	6,423.00	520.00	2.69	26.81	1.30	1.39
16.29	7.10	23.00	8.33	4.00	11.35	4.96	6,568.00	517.00	2.16	27.75	1.39	1.48
18.92	7.70	21.97	8.15	3.82	11.07	4.97	6,568.00	529.00	2.42	29.35	1.48	1.47
19.61	10.57	20.20	8.13	6.00	11.23	4.83	6,556.00	520.00	2.39	29.35	1.47	1.41
19.61	13.59	20.60	8.13	6.20	10.15	5.02	6,471.00	520.00	2.46	29.35	1.41	1.30
19.61	13.59	21.08	8.13	5.75	10.15	4.81	6,614.00	520.00	2.90	29.35	1.30	1.30
16.00	13.59	19.23	8.13	7.62	11.70	5.05	6,718.00	520.00	3.35	29.35	1.30	1.40
18.06	13.59	20.90	13.93	7.62	10.90	4.89	6,754.00	520.00	3.23	29.35	1.40	1.30
17.47	13.59	24.50	13.90	7.62	9.85	4.73	6,840.00	520.00	2.81	29.35	1.30	1.21
19.19	13.59	26.10	13.90	7.62	9.23	4.74	6,900.00	520.00	3.06	29.35	1.21	1.27
18.00	34.88	32.94	13.90	7.62	10.10	4.73	6,692.00	670.00	2.85	29.35	1.27	1.40
15.67	14.87	28.00	8.90	7.62	10.19	4.74	6,744.00	684.00	2.84	33.49	1.40	1.38
17.24	17.81	29.80	8.90	10.46	10.19	4.82	6,696.00	675.00	2.85	41.67	1.38	1.36
17.33	16.25	26.82	8.95	5.02	10.19	4.78	6,633.00	704.00	2.76	43.66	1.36	1.30
19.14	17.25	25.60	8.32	8.06	10.37	4.70	6,649.00	715.00	2.38	42.00	1.30	1.18
20.85	21.00	25.61	8.40	9.42	10.13	5.18	6,880.00	652.00	1.92	39.00	1.18	0.99
21.56	21.00	30.42	8.40	8.83	10.04	5.53	6,782.00	652.00	1.95	39.00	0.99	0.99
21.56	21.00	31.51	8.40	8.83	10.04	5.72	6,621.00	652.00	2.06	39.00	0.99	0.96
19.33	21.00	34.33	8.40	8.83	10.35	5.53	6,740.00	652.00	2.21	39.00	0.96	0.94
17.87	21.00	37.02	10.10	8.83	10.35	5.34	6,632.00	652.00	2.47	39.00	0.94	0.94
17.87	21.00	41.39	10.10	8.83	11.17	5.15	6,759.00	652.00	2.47	39.00	0.94	0.87
16.66	21.00	44.90	10.10	8.83	10.95	5.39	6,655.00	652.00	2.33	39.00	0.87	1.00

Figure 5. The sample of the database for considering the Correlation

According to Figure 5, it demonstrates a database for considering the Correlation with the Cassava including the price of agricultural products with 11 types, the process is mainly analyzing the price of cassava with the baseline of other agricultural products, which will be analyzed by the price from the previous month, in particular, to forecast the possible price in the upcoming month.

Actual	Difference	X
1.30	-	-
1.39	0.09	True
1.48	0.09	True
1.47	-0.01	False
1.41	-0.06	False
1.30	-0.11	False
1.30	0.00	False
1.40	0.10	True
1.30	-0.10	False
1.21	-0.09	False
1.27	0.06	True
1.40	0.13	True
1.38	-0.02	False
1.36	-0.02	False
1.30	-0.06	False
1.18	-0.12	False
0.99	-0.19	False
0.99	0.00	False
0.96	-0.03	False
0.94	-0.02	False
0.94	0.00	False
0.87	-0.07	False

(a) The process of convert data

Mung Bean	Shallot	Garlic	Potato	Onion	Soy	Maize	Rice	Sugarcane	Palm Kernel Oil	Coffee	Cassava
TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE
TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE
TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE
FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE
FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE
FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE
FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE
TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE
TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE
TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE
TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE
FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE
FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE
FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE
TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE

(b) The results of the data

Figure 6 The data for considering the Association Rules

It figure 6 , is the second type of data, which demonstrates the sample of the database for considering the Association Rules with the price of cassava. This sample shows that the relation between the price of Cassava and the others agricultural products from 11 types would be configured as either "TRUE" , the going up price, or "False", the going down price.

3.4 Modeling, Forecasting and Clustering

Firstly, the relationship between the price of cassava and of the 11 types of agricultural products, including the price of cassava from the previous month. This relationship will be a feature of forecasting price of cassava in the upcoming month. Significantly, all of these databases will be implied in the main feature to forecast through the processing of Correlation with the formula of the "Calculation weighed correlation", which is performed in equation (1) [18] as below;

$$\boldsymbol{C} = \frac{(\boldsymbol{X}(\boldsymbol{i})-\overline{\boldsymbol{X}})(\boldsymbol{Y}(\boldsymbol{i})-\overline{\boldsymbol{Y}})}{(\boldsymbol{n}-\mathbf{1})\boldsymbol{S_X S_Y}} \tag{1}$$

Where:

C refers to correlation coefficient.
X, *Y* refer to attributes of factors/variables.
N refers to the total number of example.
S_X, S_Y refer to standard deviation of X and Y.
i refers to the increment variable of summation.

Secondly, the association rule is required by calculating FP-Growth Algorithm, in particular, to calculate the Support and Confidence configuration. The Support will be implied with the indication of the Item set {A}, which is appeared in the below formula (2) [20].

$$\text{Item set } \{A\}$$
$$Support(A) = \frac{Frequent\ of\ A}{Total\ number\ of\ transactions} \tag{2}$$

Meanwhile, the Confidence will be implied the indication of Confidence (A => B) as below formula 3 [20].

$$Confidence\ (A \ => \ B) \ = \ \frac{Support\ (A \boldsymbol{U} \boldsymbol{B})}{Support\ (A)} \tag{3}$$

After receiving the Correlation with the rule of association, the result will be the add-on Feature for supporting the forecasting process. The analytical process, which is implied to consider the forecast process of the Cassava's price with the 6 machine learning algorithms as follows;

1. Generalized Linear Model,
2. Deep Learning,
3. Decision Tree,
4. Random Forest,
5. Gradient Boosted Trees,
6. Support Vector Machine

Thirdly, the process of calculating with each method will be analyzed with the accuracy of the train/test ratio of 60:40 following the lowest result of the Root Mean Square Error (RMSE) and the lowest result of the Mean Absolute Percentage Error (MAPE). After that, clustering of cassava price result will be categorized into the 3 groups.

1. The peak pricing period/month,
2. The average pricing period/month,
3. The lowest pricing period/month.

And algorithms are set as parameters for analysis, such as

1. Generalized linear model use regularization.
2. Deep Learning, used model use automatically
3. Decision Tree model use automatically optimize for maximal depth.
4. Random forest model use automatically optimizes for the number of trees and maximal depth.
5. Gradient Boosted Trees used model use automatically optimize for the number of trees, maximal depth, and learning rate
6. Support Vector Machine model use automatically optimize. In the meantime, other algorithms are used as a default condition.

3.5 Evaluation

The Percentage of Accuracy will be implied to evaluate with each algorithm, including the Calculation of accuracy and the Root Mean Square Error (RMSE), which comply with the standard instruction to measure avoiding the error of a calculation process in term of the quantitative approach. As a result, the process will be calculated as below formula [10] .

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2} \tag{4}$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \tag{5}$$

The MAPE is defined as the average of the absolute error expressed in percentage over a sample, or by the form where:

n: represents the number of fitted points
Y_t: represents the actual value
$\hat{Y}_t$: represents the predicted value

3.6 Deployment

The research result is expedient for those stakeholders. For instance, Agriculture, Relevant Agencies. Significantly, they will imply the study to arrange their planting plans for earning the best price for their agricultural products.

4. RESULTS

4.1 The Correlation Result

To evaluate the experiments and results, the study considers the Correlation with the Cassava including the price of agricultural products with 11 types as shown in Table 1.

Table 1. The Correlation result of Cassava including the other agricultural products with 11 types\

Number	Name	Correlation
1	Cassava	0.95
2	Maize	0.68
3	Soy	0.51
4	Sugarcane	0.44
5	Mung Bean	0.42
6	Coffee	0.38
7	Palm Kernel Oil,	0.34
8	Potato	0.25
9	Rice	0.20
10	Garlic	0.07
11	Shallot	0.07
12	Onion	0.01

According to Table 1, the study significantly founds that the Correlation of Cassava's price compared to other agricultural products is necessary to be an add-on feature in the analytical process because the result shows that the price of cassava in the previous month of the column "Cassava" is the highest result as 0.95. Therefore, the study highly requires forecasting by implying the price cassava combining with the maize to forecast the price in the upcoming month, in particular, one of the important factors to forecast has to rely on the considering the Correlation to be a feature in the analytical process of cassava's price.

4.2 The Association Rule Result

Following the study, the result finds out the association rule with FP-Growth Algorithm, this calculation is inquired "Support" and "Confidence" configuration to comply with the rule as an add-on feature, the association rule represents as below Table 2.

Table 2. The result of finding the associations rule with FR-Growth

Premises	Conclusion	Support	Confidence
Corn	cassava	0.28	0.64
Mung Bean	cassava	0.30	0.58
Soy	cassava	0.34	0.57
Sugarcane	cassava	0.32	0.54
Potato	cassava	0.41	0.54
Potato, garlic	cassava	0.30	0.52
Onion, Sugarcane	cassava	0.27	0.52
Potato, Shallot	cassava	0.27	0.52
Garlic	cassava	0.36	0.50
Shallot	cassava	0.36	0.50
Onion, Potato	cassava	0.28	0.50

In Regard to Table 2, the result of the association rule represents the significance of cassava price, which is relied on the factor of "Support" and "Confidence" configuration. Especially, the rule of Premises and Conclusion shows the rising price of Corn, which relates to the price of cassava directly; as we can review from the configuration database, the "Support" is 0.28, the "Confidence" is 0.64, as well as, choosing the "Confidence", which is higher than 0.5, it belonged to the 11 types of agricultural product implying as an add-on feature in the forecasting process.

4.3 The Forecasting Result

When the study found the Correlation configuration including the association rule as an add-on feature consequently, the forecasting process would be counted with six approaches.

1. Generalized Linear Model (GLM)
2. Deep Learning (DL)
3. Decision Tree (DT)
4. Random Forest (RF)
5. Gradient Boosted Trees(GBT)
6. Support Vector Machine (SVM)

which is set the parameter to build a model with the approach of 10 – cross Validation, the result of the process shown as below Table 3.

Table 3**.** The result of finding Root Mean Square Error (RMSE)

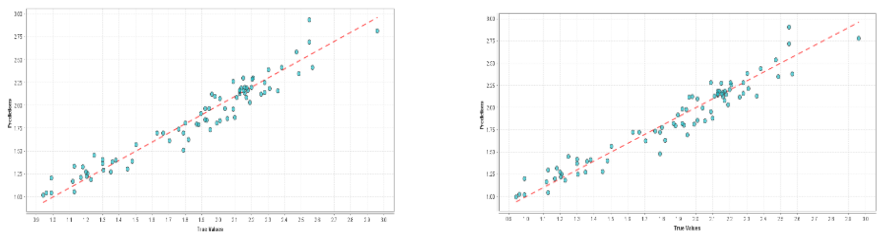
Price / Algorithms	**GLM**	**SD**	**DL**	**SD**	**DT**	**SD**	**RF**	**SD**	**GBT**	**SD**	**SVM**	**SD**
Correlation												
Cassava	0.12	0.02	0.12	0.03	0.14	0.03	0.12	0.03	0.15	0.02	0.13	0.02
Maize	0.12	0.02	0.14	0.02	0.14	0.03	0.12	0.03	0.15	0.03	0.13	0.02
Soy	0.12	0.02	0.14	0.01	0.14	0.03	0.13	0.03	0.15	0.03	0.13	0.02
Sugarcane	0.12	0.01	0.13	0.02	0.14	0.03	0.14	0.02	0.14	0.03	0.12	0.02
Mung Bean	0.12	0.02	0.12	0.03	0.14	0.03	0.15	0.02	0.16	0.02	0.12	0.02
Coffee	0.12	0.02	0.13	0.02	0.15	0.02	0.13	0.02	0.14	0.03	0.13	0.02
Palm Kernel Oil	0.12	0.02	0.13	0.02	0.14	0.03	0.12	0.03	0.14	0.03	0.13	0.02
Potato	0.12	0.02	0.14	0.02	0.14	0.03	0.12	0.03	0.14	0.04	0.12	0.02
Rice	0.12	0.02	0.14	0.03	0.14	0.03	0.14	0.03	0.14	0.03	0.12	0.02
Garlic	0.12	0.02	0.13	0.03	0.14	0.03	0.14	0.03	0.15	0.02	0.12	0.02
Shallot	0.13	0.02	0.13	0.02	0.14	0.03	0.14	0.02	0.14	0.02	**0.11**	0.01
Onion	0.12	0.02	0.12	0.03	0.14	0.03	0.12	0.03	0.14	0.03	0.13	0.02
0.1 Garlic , Shallot , Onion	**0.11**	0.03	0.12	0.02	0.14	0.03	0.13	0.03	0.14	0.02	0.12	0.02
0.2 Potato, Rice	0.12	0.02	0.13	0.02	0.14	0.03	0.14	0.03	0.14	0.04	**0.11**	0.02
0.3 Coffee, Palm Kernel Oil	0.12	0.02	0.12	0.03	0.17	0.01	0.12	0.02	0.14	0.04	0.12	0.02
0.4 Sugarcane , Mung Bean	0.12	0.01	0.13	0.03	0.14	0.03	0.13	0.03	0.15	0.02	**0.11**	0.02
0.5 Maize, Soy	0.13	0.02	0.13	0.03	0.14	0.03	0.13	0.04	0.15	0.03	0.12	0.03
Association Rule												
Potato, Onion	0.12	0.02	0.12	0.03	0.14	0.03	0.14	0.03	0.14	0.03	0.13	0.02
Potato, Shallot	0.13	0.02	0.12	0.02	0.14	0.03	0.13	0.03	0.14	0.03	0.12	0.02
Onion, Sugarcane	0.11	0.02	0.12	0.03	0.14	0.03	0.12	0.03	0.14	0.03	0.12	0.02
Garlic, Potato	0.12	0.02	0.12	0.03	0.14	0.03	0.13	0.03	0.14	0.03	**0.10**	0.02
AVG	0.12	0.02	0.13	0.02	0.14	0.03	0.13	0.03	0.14	0.03	0.12	0.02

According to Table 3, the result of finding Root Mean Square Error (RMSE) significantly represents that implying the association rule comes up with the efficiency result in the forecasting process, in particular, the effectiveness of the add-on feature is Garlic's Price. Meanwhile, Potato's Price shows the lowest Root Mean Square Error (RMSE) as 0.10 with the SVM approach. This approach also functionally works in the forecasting analysis.

The research found that the finding of configuration with the association rule as an add-on feature is better than the "Correlation" approach but the efficiency is similar.

The implication of the association rule with the add-on feature; Garlic's price and Potato's price, found that the "Support" configuration is 0.30, the "Confidence" configuration is 0.52, which is the highest from that rule, and also related to Cassava's price. Additionally, the study also found that Potato's price shows up the highest in the "Support" configuration at 0.41, and Garlic's price is 0.36 respectively. Regarding the research method, the forecasting process is not upon on the amount of database to predict because there is the various pattern with the difference, which is uncertainty.

In addition, there are 3 approaches for the forecasting process; 1) Linear Model, 2) Neural Network, and 3) Tree. The study found that the best approach to forecast is the Support Vector Machine and Generalized Linear Model, which is called as Linear Model, this approach significantly conducts a way of describing a response variable in terms of a linear combination of predictor variables. Consequently, the response should be a continuous variable and be at least approximately normally distributed. Even though each model finds wide applications, it cannot handle clearly discrete or skewed continuous responses. [22] This model is more effective than the Deep Learning, which is Neural Network, and also better than the Decision Tree, the Random Forest, and the Gradient Boosted Trees as well as the Tress structure as below details. [18]



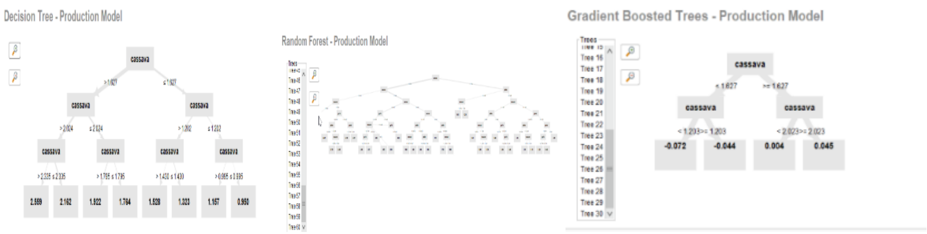
Support Vector Machine Generalized Linear Model

(a) Linear Model

Duration	Training Speed	Epochs	Iterations	Samples	Training RMSE
0.000 sec		0.00000	0	0.000000	NaN
0.016 sec	17636 obs/sec	1.00000	1	194.000000	0.15911
0.028 sec	18476 obs/sec	2.00000	2	388.000000	0.13882
0.041 sec	18774 obs/sec	3.00000	3	582.000000	0.15177
0.052 sec	19400 obs/sec	4.00000	4	776.000000	0.13559
0.064 sec	19795 obs/sec	5.00000	5	970.000000	0.15360
0.076 sec	19728 obs/sec	6.00000	6	1164.000000	0.13821
0.090 sec	19126 obs/sec	7.00000	7	1358.000000	0.14918
0.102 sec	19400 obs/sec	8.00000	8	1552.000000	0.17098
0.113 sec	19617 obs/sec	9.00000	9	1746.000000	0.14071
0.124 sec	19795 obs/sec	10.00000	10	1940.000000	0.13338

Deep Leaning

(b) Neural Network



Decision Tree Random Forest Gradient Boosted Trees

(c) Tree

Figure 7. The structure of the six algorithms (a) Linear Model, (b) Neural Network (c) Tree

4.4 The Error Forecasting

There are 2 types of error in the forecasting process; the higher price from the actual and the lower price

Table 4. The Errors in the forecasting process.

Time	Real price	Forecasting	Difference
Dec 2009	1.79	1.51	0.28
Sep 2010	2.55	2.93	-0.38
Aug 2016	1.25	1.45	-0.20
Apr 2019	2.11	1.86	0.24

In Regard to the Error configuration as of December 2019, the forecasting process is lower than the actual, the study found that the error result of the Cassava's price in November 2009 is 1.45 THB, and the price for December 2019 is 1.79 THB. The price fluctuation during those periods (1.45 – 1.79) is 0.34 THB.

Therefore, the forecasting process comes up with a lower price than the actual of 0.280. Meanwhile, the price of Garlic and Potato are unchanged at 72 THB and 13.70 THB, which present a similar price from November – to December 2009. In addition, the error configuration as of September 2010, August 2016, and April 2019 also share the same direction.

According to the results of table 3, the study found that the best method regarding 6 algorithms is the Support Vector Machine approach, which is main analysis in the forecasting process and it doesn't, which is implied with the association rule as an add-on feature represent the lowest Root Mean Square Error (RMSE) as 0.10. After that, this study compares to the Conventional Method; the result shows that the proposed method demonstrates the value of the Mean Absolute Percentage Error (MAPE) as 3.35%, it displays more effective as 0.61%, shown as table 5.

Table 5. The result shows compare to the Conventional Method

Method	Tool	MAPE(%)
Conventional Method (16)	Multi-Layer Artificial Neural Network	3.96%
Proposed Method	Support Vector Machine	3.35%

After that, the forecasting process had been analyzed from March 2021 – to February 2022 (12 months), the results shown in Table 6.

Table 6. The result of the forecasting

Month-Year	Real Price	Forecasting
Mar-21	2.13	2.07
Apr-21	2.08	2.09
May-21	1.92	2.03
Jun-21	1.91	1.88
Jul-21	1.97	1.87
Aug-21	2.02	1.93
Sep-21	2.12	1.98
Oct-21	2.08	2.08
Nov-21	2.19	2.04
Dec-21	2.28	2.15
Jan-22	2.29	2.23
Feb-22	2.28	2.24

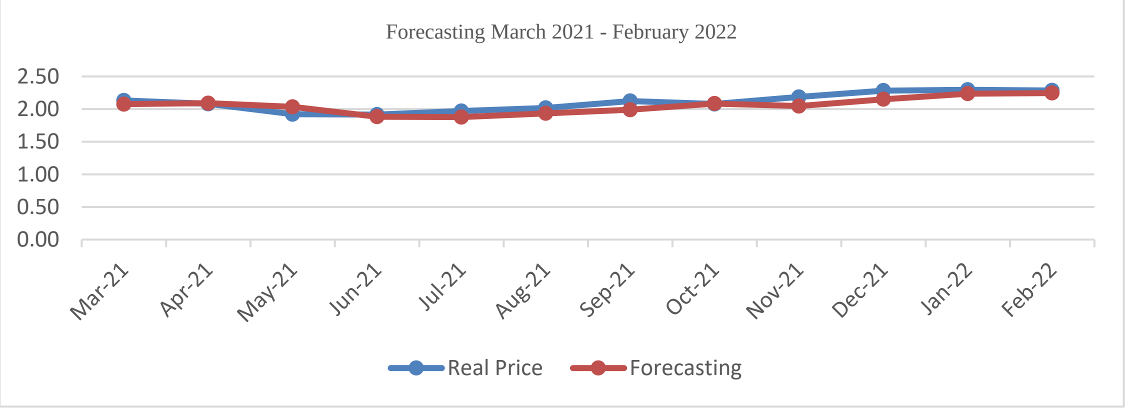


Figure 8. Forecasting

The RMSE configuration is 0.086 and The MAPE is 3.29% according to the result of forecasting in Table 6. and the forecasting from Figure 8. The level of analysis presents the highly accurate forecasting if compare to the MAPE criteria for model evaluation [16].

Table 7. MAPE criteria for model evaluation

MAPE (%)	Interpretation
<10	Highly accurate forecasting
10–20 Good forecasting	10–20 Good forecasting
20–50 Reasonable forecasting	20–50 Reasonable forecasting
>50 Inaccurate forecasting	>50 Inaccurate forecasting

Lastly, the clustering of Cassava's price is the last process of the study, it provides clustering the pricing into a group yearly, which is implied by k-mean. There are 3 groups of the Cassava prices: 1. Peak sales period 2. The average sales period and 3. The minimum sales period, the result shown Table 8. below

Table 8. The clustering of Cassava's price **by monthly**

Month	Peak pricing period (times)	percentage	Average pricing period (times)	Percentage	Minimum pricing period (times)	percentage
January	8	11.26	3	5.66	5	7.69
February	8	11.26	3	5.66	5	7.69
March	8	11.26	4	7.54	4	6.15
April	7	9.85	4	7.54	4	6.15
May	6	8.45	4	7.54	5	7.69
June	3	4.22	6	11.3	6	9.23
July	2	2.81	7	13.21	7	10.76
August	6	8.45	3	5.66	7	10.76
September	5	7.04	4	7.54	7	10.76
October	3	4.22	4	7.54	9	13.84
November	5	7.04	7	13.21	4	6.15
December	10	14.08	4	7.54	2	3.07

Regarding Table 8, the result of the clustering of Cassava's price, the study found that the peak pricing period in December is 14.08%. The average pricing period is in July, and the minimum pricing period is in October. At this point, the best period for selling cassava is December, which is the highest price. The price presents the highest period twice times. On the other hand, the worst period for selling cassava is July and December. The price presents the lowest rate 7-9 times. This clustering process aims to present the best period of Cassava's price.

5. DISCUSSION

According to the forecasting Cassava's price, the authors found that the other study also implied with the feature or variable to support the forecast process. Meanwhile, the forecast process in this study is analyzed by the Correlation and Association rule as an add-on feature. In details, the authors apply the analytical method in the forecasting process by bringing the price of agricultural products compare with the Cassava's price. As a result, the highest result as 1 represent the configuration of the Correlation and Association Rule following its result. For instance, if Potato's Price and Garlic's price are increased or decreased, Cassava's price will be following to those agricultural prices. In addition, the authors also considered this approach to implied as an add-on feature, which is a crucial variable to support the forecasting process for Cassava's price with the 6 methods; (1) Support Vector Machine, (2) Generalized Linear Model, (3) Deep Learning, (4) Decision Tree, (5) Random Forest, and (6) Gradient Boosted Trees. Technically, there are 3 approaches for the forecasting process;

1) Linear Model such as Support Vector Machine, Generalized Linear Model
2) Neural Network such as Deep Learning
3) Tree such as Decision Tree, Random Forest, Gradient Boosted Trees

Finally, the study found that the Linear Model; Support Vector Machine and Generalized Linear Model is the best method for forecasting process. The following method is Random Forest, Decision Tree, Gradient Boosted Tree, and the least is Neural Network, Deep Learning. The study also found that Support

Vector Machine is used for time series forecasting which is related to the outcomes of other studies [26], [27]. For Deep Learning, this approach is suitable for large data set such as, Big Data. However, it is not a suitable approach for this study, which presents with the minimum of data set. For Error in forecasting process, the study also found the high vulnerable of Cassava's price, this also make the forecasting process discrepancy. The final approach is the clustering of Cassava's price, which is catetagorized into the 3 groups: (1) Peak sales period (2) The average sales period, and (3) The minimum sales period. In details, this step also involved with k-mean for clustering the Cassava's price, which is also worked effectively.

6. CONCLUSION

The study found that the forecasting process would reach the maximum efficiency, if bringing, it implies forecasting the Cassava's price. However, the study also found that fluctuating price is the main obstacle to forecasting. A further suggestion is to find a suitable feature or configuring the parameters for the forecasting analysis. The instruction would be improved the research progression. Subsequently, the agricultures could apply the research to implement their planting plans for profit-making.

ACKNOWLEDGEMENTS

The researchers put a special thanks to the Office of Agricultural Economics, Ministry of Agriculture, and Cooperatives in supporting secondary data used to analyze in the study.

REFERENCES

[1] Bank of Ayudhya Public Company Limited, "Cassava Industry", 2018, from https://www.krungsri.com/enresearch/industry/industry-outlook/agriculture/cassava/IO/io-cassava-21

[2] B. Nirasok, "Analysis of the costs and benefits of planting cassava in the central area,", *The 1st Rajamangala University of Technology Suvarnabhumi National Conference, Vol 1,* pp. 987-993, 2016.

[3] S. Deepradit, *et al.*, "The Study of Forecasting Techniques for Aromatic Coconut Monthly Prices Using Individual and Hierarchical Forecasting,", *Thai Journal of Operations Research* , Vol 2, pp. 16-26, 2020.

[4] S. Deepradit and P. Ruksorn, "The Study of Forecasting Techniques for Aromatic Coconut Monthly Prices Using Individual and Hierarchical Forecasting,", *Thai Journal of Operations Research* , Vol 2, pp. 16-26, 2020.

[5] S. Khamrod, *et al.*, " Price Forecasting Model of Litopenaeus Vannamei size of 70 counts per Kilogram using Box-Jenkins Method,", *Association Of Private Higher Education Institutions Of Thailand,* pp. 34-44, 2019.

[6] B. Choopradit and S.Chaipitak., "Forecasting model for mango export volumes of Thailand,", *Naresuan University Journal: Science and Technology* ,Vol 26 , pp. 74-85, 2018.

[7] N.Thitinunpong and P.Parthanadee, " Forecasting of Fresh Cassava Root Buying Prices and Cassava Chip Selling Prices,", *The 50th KU Annual Conference,* pp. 26-33, 2012.

[8] W.Keerativibool, " Forecasting Model for the Export Volumes of Cassava ,", *Thaksin University Journal,vol 2,* pp. 31-43, Vol 2, 2016.

[9] B.R. Kwasi and B.J. Kobina, " Forcasting Of Cassava Prices In The Central Region Of Ghana Using Arima Model", *Intercontinental Journal Of Marketing Research Review,* pp. 26-33, Vol 2, 2014.

[10] R. Ponprasert and S. Tepdang, " Forecasting of cassava price by using Box-Jenkins method," , *Rattanakosin Journal of Science and Technolog,* vol 4. 10, pp. 22-33, 2022.

[11] S.Shen ,*et al.*, "Stock Market Forecasting Using Machine Learning Algorithms," , *Department of Electrical Engineering, Stanford University, Stanford, CA,* 2012.

[12] B.M. Pavlyshenko, "Machine-Learning Models for Sales Time Series Forecasting , *MMDPI,* 2019.

[13] F. Kamalov,*et al.*, "Financial Forecasting with Machine Learning: Price Vs Return," , *Journal of Computer Science,* vol 17(3), 251-264, 2021.

[14] J. K. Mbugua,*et al.*, "Predicting Suitable Areas For Growing Cassava Using Remote Sensing And Machine Learning Techniques: A Study In Nakhon-Phanom Thailand," , *Issues in Informing Science and Information Technology,* vol 15, 43-56, 2018.

[15] O. Olanloye and E. Oduntan, " Machine Learning Predictive Modeling Of The Price Of Cassava Derivative (Garri) In The South West Of Nigeria," , *Applied Computer Science,* vol. 14, pp. 53–63 , 2019

[16] P.Boonrawd and K.Phonyiam. " A Cassava Price Forecasting Model Using aMulti-Layer Artificial Neural Network," , *Thai Science and Technology Journal,* vol. 3, pp. 533-543, 2017.

[17] E. Pacharawongsakda, *An Introduction to Data Mining Techniques*, first ed. Bangkok, Asia Digital Press Co., Ltd, 2014.

[18] S. Tepdang and R. Pongprasert, "Forecast of Changes in Exchange Rate between Thai Baht and US Dollar Using Data Mining Technique" , *S NRU Journal of Science an d Technology,* vol. 12(3), pp. 213 – 221, 2020.

[19] The Office of Agricultural Economics, Ministry of Agriculture, and Cooperatives."Agricultural Price Index",2022, from https://www.oae.go.th/view/1/%E0%B8%94%E0%B8%B1%E0%B8%8A%E0%B8%99%E0%B8%B5%E0%B8%A3%E0%B8%B2%E0%B8%84%E0%B8%B2%E0%B9%81%E0%B8%A5%E0%B8%B0%E0%B8%9C%E0%B8%A5%E0%B8%9C%E0%B8%A5%E0%B8%B4%E0%B8%95/TH-TH

[20] The Office of Agricultural Economics, Ministry of Agriculture, and Cooperatives."Agricultural Price Index" 2022, from https://www.oae.go.th/view/1/%E0%B8%94%E0%B8%B1%E0%B8%8A%E0%B8%99%E0%B8%B5%E0%B8%A3%E0%B8%B2%E0%B8%84%E0%B8%B2%E0%B9%81%E0%B8%A5%E0%B8%B0%E0%B8%9C%E0%B8%A5%E0%B8%9C%E0%B8%A5%E0%B8%B4%E0%B8%95/TH-TH

[21] S. Kanwaseth , Ni-Universit. "Association Rules", 2022. from https://medium.com/tni-university/association-rules-a36515861b6e

[22] sciencedirect,"Linear models", 2022. from https://www.sciencedirect.com/topics/mathematics/linear-models#:~:text=Linear%20models%20are%20a%20way,discrete%20or%20skewed%20continuous%20responses.

[23] L.Hsu and C.Wang, "Applied multivariate forecasting model to tourism industry," , *An international Interdisciplinary Journal*, 2008.

[24] S. Tepdang, T. Choosang, and S. Sothanakul, “The Association and Prediction of Index of the Stock Exchange of Thailand (SET) between the Index of Foreign Exchange by using Data Mining Techniques”. IEICE Proceedings Series, 61(W1-1-3), 2016.

[25] M. R Pahlawan, E. Riksakomara, R. Tyasnurita, A., Muklason, F. Mahananto, and R. A. Vinarti. “Stock price forecast of macro-economic factor using recurrent neural network”. IAES International Journal of Artificial Intelligence, 10(1), 74. 2021.

[26] R. Samsudin, A. Shabri and P. Saad, "A Comparison of Time Series Forecasting using Support Vector Machine and Artificial Neural Network Model", *Journal of Applied Sciences*, , vol. 10(11), pp. 950 – 958, 2010.

[27] H. Kibtia, S. Abdullah, and A. Bustamam. Comparison of random forest and support vector machine for prediction of cognitive impairment in Parkinson's disease. ", *AIP Conference Proceedings*, 2020.

BIOGRAPHIES OF AUTHORS

Sayan Tepdang received his B.Sc. and M.Sc. (Computer science) from King Mongkut's University of Technology North Bangkok and Thammasat University, Thailand in 2005 and 2010, respectively. He is currently lecturer at Faculty of Business Administration and Information Technology, Rajamangala University of Technology Tawan-Ok, Chakrabongse Bhuvanarth Campus, Bangkok, Thailand. He researches areas of interest include Artificial Intelligent, Natural Language Processing, Computer Vistion and Data Minging. He can be contacted at email: sayan_te@rmutto.ac.th

Ratthakorn Ponprasert received the Bachelor of Economics (B. Econ.) and Master of Business Economics (M. Econ.) from Mahasarakham University, Thailand in 2012 and 2015, respectively. He is currently lecturer at Faculty of Business Administration and Information Technology, Rajamangala University of Technology Tawan-Ok, Chakrabongse Bhuvanarth Campus, Bangkok, Thailand.He research areas of interest include Data Minging, Forecasting, and mathematical model. He can be contacted at email: ratthakorn_po@rmutto.ac.th